Bounded expansion in models of webgraphs

Silvia Gago [*][†] Dirk Schlatter [‡][†]

Abstract

We study the bounded expansion of several models of web graphs. We show that various deterministic graph models for large complex networks have constant bounded expansion. We study two random models of webgraphs, showing that the model of Bonato has not bounded expansion, and we conjecture that the classical model of Barabási may have also not bounded expansion.

1 Introduction

For the past decade, there has been a growing interest in finding suitable models for various real-world networks which share some characteristics like a small diameter or a power-law degree distribution. For surveys on the subject we refer to [4,6]. Some of these models are defined deterministically and others use randomness, but they all produce *sparse* graphs, i.e. graphs where the number of edges is only linear in the number of vertices. It is therefore natural to ask whether these graphs fall into some of the well-known sparse graph families like minor-closed graph families or families of graphs with bounded degree.

Nešetřil and Ossona de Mendez [12–15] generalized such families by defining a sequence of graph parameters. Let $\mathcal{P} = \{V_1, \dots, V_p\}$ a family of balls of G, that is, a subset of vertices inducing a connected subgraph. The set of all the families of balls of G is noted by $\mathcal{B}(G)$. The radius $\rho(\mathcal{P})$ of $\mathcal{P}$ is

[*]Universitat Politècnica de Catalunya, Department de Matemàtica Aplicada IV, Avda. Canal Olímpic s/n, 08860 Castelldefels, Barcelona, Spain, sgago@ma4.upc.edu

[†]Charles University, Institute for Theoretical Computer Science and Department for Applied Mathematics, Malostranské nám. 25, 118 00 Praha, Czech Republic, research supported by the European project AEOLUS

[‡]Humboldt University Berlin, Institute for Computer Science, Unter den Linden 6, 10099 Berlin, Germany, schlatter@informatik.hu-berlin.de

$\rho(\mathcal{P}) = \max_{X\in\mathcal{P}} \rho(G[X])$, then the parameter $\nabla_r(G)$ is defined in [12] as the *greatest reduced average density (grad)* of rank r of the graph G

$$\nabla_r(G) = \max_{\substack{\mathcal{P} \in \mathcal{B}(G) \\ \rho(\mathcal{P}) \leq r}} \frac{\mid E(G/\mathcal{P}) \mid}{\mid \mathcal{P} \mid}.$$

If this sequence $(\nabla_r(G))_{r\geq 0}$ has a uniform upper bound $f(r)$ for all graphs in a certain family, then we say that this family has *bounded expansion*, that is, a class of graphs $\mathcal{C}$ has bounded expansion if there exists a function $f : \mathbb{N} \to \mathbb{R}$ such that for every graph $G \in \mathcal{C}$ and every r holds

$$\nabla_r(G) \leq f(r).$$

2 Deterministic models

We are going to use the following results of [12, 15] to prove that the deterministic models have constant bounded expansion.

Theorem 2.1. *For any proper minor closed class of graphs $\mathcal{K}$, and for any fixed integer $p \geq 1$, $\chi_p(G)$ is bounded on $\mathcal{K}$.*

We also consider Theorem 8.1. of [12] which holds that for a class of graph $\mathcal{C}$ is equivalent to have bounded expansion than the fact that for any integer p, $\{\chi_p(G) : G \in \mathcal{C}\}$ may be bounded. Combining both theorems we have that proper minor classes have bounded expansion.

Recursive clique-trees
For some integer $d \geq 2$, recursive d-clique trees are constructed as follows: starting with $G_0 := K_d$, we obtain G_{t+1} from G_t by adding a new vertex for each clique of size d in G_t and joining this vertex to all vertices in the respective clique. The case $d = 2$ has been considered in [9] and the general case in [8]. We will denote the family of recursive d-clique trees by $\mathcal{P}_d$. If we introduce only a single new vertex in each step (and connect it to all vertices in some d-clique), we obviously get a larger graph family $\mathcal{P}'_d$ which contains $\mathcal{P}_d$. Denote the closure of $\mathcal{P}'_d$ under taking subgraphs by $\mathcal{P}''_d$. Our first result implies that recursive clique-trees have constant expansion.

Proposition 2.2. *$\mathcal{P}''_d$ is a proper minor-closed graph family.*

The construction of a d-dimensional Apollonian network is very similar to the construction of a recursive d-clique tree, only that we now introduce new vertices for each clique of size d in G_t *which does not already lie in* G_{t-1}. Therefore, the family $\mathcal{Q}_d$ of all such networks is a subset of $\mathcal{P}_d$, so that our previous result implies that $\mathcal{Q}_d$ also has constant expansion.

For some integer $d \geq 2$, recursive d-clique trees are constructed as follows: Starting with $G_0 := K_d$, we obtain G_{t+1} from G_t by introducing a new vertex for each clique of size d in G_t and joining this vertex to all vertices in the respective clique. The case $d = 2$ has been considered in [9] and the general case in [8].

We will denote the family of recursive d-clique trees by $\mathcal{P}_d$ and the closure of $\mathcal{P}_d$ under taking subgraphs by $\mathcal{P}'_d$.

Lemma 2.3. *$\mathcal{P}'_d$ is a proper minor-closed graph family.*

Proof. From the inductive construction of G_t above, it can never contain K_{d+2} as a subgraph, and $\mathcal{P}'_d$ is thus a proper graph family. If H is a minor of G, then it is well-known that we can get H from G by a sequence of edge deletions and contractions. Obviously, edge deletions will not get us out of $\mathcal{P}'_d$. To complete the proof, it is therefore sufficient to prove that for an edge $e = uv$ in a graph $G \in \mathcal{P}'_d$, G/e also lies in $\mathcal{P}'_d$.

Let t' be such that $G \subseteq G_{t'}$. We assign to every vertex x in $G_{t'}$ a *birthtime* t_x which is the smallest integer such that $x \in V(G_{t_x})$. We say that a vertex y is a *child* of a vertex x if $xy \in E(G_{t_y})$, i.e. x lies in the clique of size d which gives rise to y. We define the terms *ancestor* and *descendant* accordingly. Note that a child is always adjacent to its parent, where as a descendant of a later generation may well not be.

Using the symmetry of $G_1 = K_{d+1}$, we may assume without loss of generality that $t_u < t_v$. Let us make two observations. First, let Y be the set of descendants of y which are . $\square$

Apollonian networks

The construction of a d-dimensional Apollonian network [16] is very similar to the construction of a recursive d-clique tree, only that we now introduce new vertices for each clique of size d in G_t *which does not already lie in* G_{t-1}. Therefore, the family $\mathcal{Q}_d$ of all such networks is a subset of $\mathcal{P}_d$, so that our previous result implies that $\mathcal{Q}_d$ also has constant expansion.

Hierarchical networks

This model was introduced in [2]. Again, we start with a K_d. We select a root r in K_d and N_0 be the set of non-root vertices. We construct G_{t+1} from G_t as follows: add $d-1$ disjoint copies $G_t^{(1)}, \ldots, G_t^{(d-1)}$ of G_t to G_t and connect all the vertices in $\bigcup_{i=1}^{d-1} N_t^{(i)}$ to r. Finally, set $N_{t+1} := N_t \cup \bigcup_{i=1}^{d-1} N_t^{(i)}$. We denote the family of all such hierachical networks by $\mathcal{R}_d$.

If we are more generous and connect *all* vertices in the copies of G_t to r, we get another graph family $\mathcal{S}_d$. We denote the closure of $\mathcal{S}_d$ under taking subgraphs by $\mathcal{S}'_d$. It is then easy to see that $\mathcal{S}'_d$ is a proper minor-closed graph family and as $\mathcal{R}_d \subset \mathcal{S}'_d$, this proves that $\mathcal{R}_d$ also has constant expansion.

Sierpinski grids

A Sierpinski grid is a subset of the 2-dimensional grid which has some self-similarity [3]. An additional vertex is added which is connected to vertices in on an 'inner' boundary and vertices on the outer boundary. It is straightforward to see that this graph is embeddable on a torus and thus has constant expansion.

3 Random networks

If we want to discuss whether *random* models also have bounded expansion, we first have to clarify what we mean by that. For a given random model, we could of course ask whether the family of all graphs which might occur as the outcome, i.e. which have positive probability, has bounded expansion. However, this approach somehow seems to neglect the randomness and in most cases it trivially leads to the result that the expansion is unbounded. Instead, we will adopt the following

Definition 3.1. Let $(G_t)_{t\geq 0}$ be a random graph process, we say that it has *bounded expansion* if there exist a function $f : \mathbb{N} \longrightarrow \mathbb{N}$ such that for all $r \in \mathbb{N}$

$$\mathbb{P}\left[\nabla_r(G_t) \leq f(r)\right] \longrightarrow 1 \quad \text{as} \quad t \longrightarrow \infty.$$

In this section, instead of working with the definition of $\nabla_r(G)$ itself, we will use a result by Dvořák [10], namely that $\nabla_r(G)$ is large if and only if G contains a $\leq 2r$-subdivision of a graph with large minimum degree.

We will discuss the two most prominent random models for complex networks, the copying model of Kumar et al. [11] and the preferential attachment model of Barabási and Albert [1].

Copying model

In the copying model, we define a random graph process $(G_t)_{t\geq 0}$ inductively as follows.

1. Start with a graph G_0 consisting of two vertices v_1, v_2 and the edge $\{v_1, v_2\}$.

2. Given G_t, choose a vertex $v \in V(G_t)$ uniformly at random. Add a new vertex v_{t+1} and join it to each neighbour of v (independently) with some (fixed) probability $0 < p < 1$.

It is straightforward to prove that

Proposition 3.2. *For all $r \in \mathbb{N}$,*

$$\mathbb{P}\left[G_t \textit{ contains}\, K_{r,r}\right] \longrightarrow 1 \quad \textit{as} \quad t \longrightarrow \infty.$$

We note that this also follows from a result in [7]. It obviously implies that the copying model has unbounded expansion.

Preferential attachment model

For the preferential attachment model, we adopt the following rigorous definition of Bollobás and Riordan [5]. First we generate a sequence of graphs $(G_t^1)_{t\geq 0}$ as follows.

1. Start with a graph G_0^1 consisting of a single vertex v_1 and the loop $\{v_1, v_1\}$.

2. Given G_t^1, add a new vertex v_{t+1} and an edge $\{v_{t+1}, v\}$, where v is chosen randomly with

$$\mathbb{P}\left[v = v_i\right] = \begin{cases} d_{G_t^1}(v_i)/(2t-1) & \text{if } 1 \leq i \leq t, \\ 1/(2t-1) & \text{if } i = t+1. \end{cases}$$

To get $(G_t^m)_{t\geq 0}$ from $(G_t^1)_{t\geq 0}$ for some fixed integer $m \geq 2$, we take those graphs from the latter sequence for which m divides t and contract $v_1, \ldots, v_m$ to a new node v_1^m, $v_{m+1}, \ldots, v_{2m}$ to a new node v_2^m, and so forth.

In this model, it is not at all obvious whether G_t^m is likely to contain a $\leq 2r$-subdivision of K_r, say. Using ideas and techniques from [5], we strongly conjecture the following.

Conjecture 3.3. *Let H be a 1-subdivision of K_r. For all $r \in \mathbb{N}$,*

$$\mathbb{P}\left[G_t \textit{ contains a copy of}\, H\right] \longrightarrow 1 \quad \textit{as} \quad t \longrightarrow \infty.$$

'Proof'. We will attempt to prove that there are a.a.s. $R := \binom{r}{2}$ vertices $v_1, \ldots, v_R$ in $(\sqrt{n}, n]$ such that $G_t^m[1 \cup \ldots \cup r \cup v_1 \cup \ldots \cup v_R]$ contains a copy of H (with $1, \ldots, r$ corresponding to the vertices in H of degree $r-1$).

To this end we will use the following result of Bollobás [4], where we direct an edge $\{i, j\}$ of G_1^n from j to i whenever $i < j$. For a subgraph S of G_1^n, $d_S^{\text{in}}(.)$, $V^-(S)$, and $V^+(S)$ then have their obvious meaning; whereas $C_S(i)$ denotes the number of edges (j, k) with $j \leq i \leq k$.

Theorem 3.4. *Let S be a possible subgraph of G_1^n of fixed order. Then*

$$\mathbb{P}\left[S \subset G_1^n\right] = \prod_{i \in V^-(S)} d_S^{\text{in}}(i)! \prod_{i \in V^+(S)} \frac{1}{2i-1} \prod_{i \notin V^+(S)} \left(1 + \frac{C_S(i)}{2t-1}\right).$$

Furthermore,

$$\mathbb{P}\left[S \subset G_1^n\right] = \prod_{i \in V^-(S)} d_S^{\text{in}}(i)! \prod_{ij \in E(S)} \frac{1}{2\sqrt{ij}} \exp\left(O\left(\sum_{i \in V(S)} C_S(i)^2/i\right)\right). \quad (1)$$

For any copy of H in G_m^n as above, we now consider some subgraph H' of G_1^{mn} such that contraction we described above would transform H' into H. Let us use 1 to calculate $\mathbb{P}\left[H' \subset G_1^{mn}\right]$. Clearly, the only vertices of H' with positive indegree are those which corresponds to the 'fat' vertices $1, \dots, r$ in H. Moreover, as all $v_i > \sqrt{n}$, the contribution of the 'slim' vertices to the sum in the exponent of the last product is negligible. Therefore, we deduce that there is a constant c such that *independent of our actual choice of* $v_1, \dots, v_R$,

$$\mathbb{P}\left[H' \subset G_1^{mn}\right] = \frac{c}{v_1 \cdot \ldots \cdot v_R}.$$

Let us denote the indicator variable of this event by X_i and consider the sum $X := \sum X_i$ over all possible choices of $v_1, \dots, v_R$. Then

$$\mathbb{E}\left[X\right] = \sum_{\sqrt{n} < v_1, \dots, v_R \le n} \frac{c}{v_1 \cdot \ldots \cdot v_R} = \Theta\left((\log n)^R\right).$$

Hence, the expected number of 1-subdivisions of K_r is unbounded. $\square$

Unfortunately, we cannot use the second moment method to actually show that we have such a subdivision with high probability. The reason is as follows. We would need to show that $\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \longrightarrow 1$, but the dominant contribution in the numerator arises from pairs $X_i X_j$ where the corresponding sets of subdividing vertices are disjoint. When we use Equation (**??**) to calculate $\mathbb{E}\left[X_i X_j\right]$, we will get a different constant c'. It can be easily seen that $c' > c^2$, so that $\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \not\longrightarrow 1$. Therefore, we can infer that also the preferential attachment model has unbounded expansion.

Our results raise the question whether there are 'natural' deterministic and random models for complex networks which have unbounded or bounded expansion, respectively.

Acknowledgment

The authors thanks to Prof. J. Nešetřil for his thoughtful comments and suggestions. This research was supported by the European project AEOLUS.

References

[1] A.-L.Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286**, 509-512, 1999.

[2] A.-L.Barabási and E. Ravasz, *Hierarchical organization in complex networks*, Phys. Rev. E **67**, (2003), 026112.

[3] L. Barriere, F. Comellas, C. Dalfo. *Fractality and the small-world effect in Sierpinski graphs.* J. of Physics A- Mathematical and general, **39** (2006), 11739-11753.

[4] B. Bollobás and O. Riordan, *Mathematical results on scale-free random graphs*, in: Handbook of graphs and networks: from the genome to the internet (eds. S. Bornholdt and H.G. Schuster), 1-34, Wiley-VCH, 2002.

[5] B. Bollobás and O. Riordan, *The diameter of a scale-free random graph*, Combinatorica **4**, 5-34, 2004.

[6] A. Bonato, *A survey of models of the web graph*, Proceedings of Combinatorial and Algorithmic Aspects of Networking, 2004.

[7] A. Bonato and J. Janssen, *Infinite limits of copying models of the web graph*, Internet Mathematics **1**, 193-213, 2004.

[8] F. Comellas, G. Fertin, and A. Raspaud, *Recursive graphs with small-world scale-free properties*, Phys. Rev. E **69**, 037104, 2004.

[9] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, *Pseudofractal scale-fee web*, Phys. Rev. E **65**, 066122, 2002.

[10] S. Dvořák, *Asymptotical structures of combinatorial objects*, PhD thesis, Charles University Prague, 2007.

[11] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, *Stochastic models for the web graph*, Proceedings of the 41st IEEE Symp. on Foundations of Computer Science, 57-65, 2000.

[12] J. Nešetřil and P. Ossona de Mendez, *Grad and Classes with Bounded Expansion I. Decompositions*, KAM-DIMATIA Series **2005-739**, 2005.

[13] J. Nešetřil and P. Ossona de Mendez, *Grad and Classes with Bounded Expansion II. Algorithmic Aspects*, KAM-DIMATIA Series **2005-740**, 2005.

[14] J. Nešetřil and P. Ossona de Mendez, *Grad and Classes with Bounded Expansion III. Restricted Dualities*, KAM-DIMATIA Series **2005-741**, 2005.

[15] J. Nešetřil and P. Ossona de Mendez, *Tree depth, subgraph coloring and homomorphism bounds*, KAM-DIMATIA Series **2004-656**, 2004.

[16] Z. Zhang, L. Rong, F. Comellas, G. Fertin, *High dimensional Apollonian networks*, J. Phys. A: Math. Gen. Vol. **39** (2006), 1811-1818.